

An Audit on Bing’s China Censorship or, an Independent Transparency Report

Xia Chu*

Version 1.1, March 1, 2014

1 Abstract

Microsoft has long been known of practicing censorship in China with its Bing search engine, but little is known on how Bing’s China-related filtering actually works. In this thorough study, we examined Bing’s SERP (search engine result page) for a large body of (30,000+) sensitive and non-sensitive query terms, queried from inside China and outside China. Comparing and examining these results, plus querying with special search operators, surfaced unprecedented details on Bing’s China filtering practice. In a sense, this is an independent “Transparency Report”.

This study is motivated by the scandal in which Microsoft was accused of practicing China censorship globally, first broken by GreatFire.org [3] in February, but denied by Microsoft [4].

Main findings from this research:

- Bing has a list of “forbidden” terms that no results are shown. 139 such terms are identified in this study.
- Bing has a blacklist of websites that it never shows to China users. 329 such websites are identified in this study.
- Bing has a huge blacklist of URLs that it never shows to China users. 1593 URLs are identified in this study, which is only the tip of the iceberg.
- “Censorship leakage”, we observed censorship notices for 1710 unique query terms on the Chinese version of Bing from outside China.

In particular, I’d like to point out two issues.

- Bing’s filtering is very broad, for example, Bing filters the entire site for five major editions of Wikipedia: Japanese, French, German, Dutch and Swedish.
- Bing has extensive filtering rules for domestic sites like blog.163.com, baidu.com, blog.sina.com.cn, 360doc.com, ifeng.com, douban.com etc. The content filtered by Bing still exists on these domestic sites and are not even filtered by domestic search engines like Baidu.

Overall, from these findings, my conclusion is that Bing has not lived up to their promise of “support for freedom of access to information” (Bing’s claims can be found in Appendix A). For example, removing the entire sites of ja.wikipedia.org, fr.wikipedia.org, de.wikipedia.org, nl.wikipedia.org and sv.wikipedia.org does not support Bing’s claim of “if we are required to implement the request, we will do so narrowly”. Also, the fact that Bing filters vast amount of domestic content and *more than Baidu does* is utterly saddening and inexcusable.

Link to this report, actual lists identified, and future updates can all be found at the master spreadsheet <http://goo.gl/qzAaVe> [1].

Update: In Version 1.0 of this paper, we listed 149 terms in the Forbidden Words list. However, it appears that Microsoft has been actively modifying its censorship system. Between Feb 20 and Feb 28, 10 terms in this list were removed, including 3 of the 5 “peculiar” terms as explained in Section 4.2.

Note: when not specified otherwise, we are always focusing on Bing’s mainland China edition. This edition is not limited to mainland Chinese users though. This setting is tied to Bing’s interface language and presumably many Chinese users abroad choose this edition for language reason (for example, those who set their language preference to zh-CN for their Windows PC).

*Author contact: Email: summer.agony@gmail.com; Twitter: @SummerAgony.

1.1 Report Structure

- Section 1: The abstract.
- Section 2: The motivation and background for this study.
- Section 3: The methodology.
- Section 4: The details of Bing censorship.
- Section 5: A deep-dive on Bing's filtering rules for New York Times.
- Section 6: A deep-dive on Bing's filtering rules for blog.163.com (a popular domestic blogging site).
- Section 7: Discuss whether Bing is performing China censorship outside China, and the potential gray censorship.
- Appendix A: Bing's claims regarding "address laws specific to individual countries".
- Appendix B: List of identified forbidden terms.
- Appendix C: List of identified sunken websites.
- Appendix D: List of identified vanished URLs.
- Appendix E: A sample of query terms which trigger censorship notice *outside China*.

2 Motivation and Background

2.1 Motivation

On Feb 11, 2014, China censorship watchdog GreatFire.org accused Bing as deploying Chinese censorship on global scale [3]. It received lots of media coverage. Soon after, Microsoft posted a response "Setting the Record Straight" denying the accusation on its search blog [4].

There are a lot of discussions on the issue, most noteworthy are Jason Q. Ng's article [5] "Reputation Matters: Unpacking the Microsoft China Censorship Scandal" (link) on Wall Street Journal and Rebecca MacKinnon's article [6] "Where is Microsoft Bing's transparency report" (link) on the Guardian.

MacKinnon called for Microsoft to "publish a Transparency Report with data on the number of government censorship requests". This is unlikely to happen any time soon, especially for China. So, I decide to take upon it - analyzing Bing's China censorship and publishing an extensive "Transparency Report" for Bing.

2.2 Different Censorship Schemes

China's internet censorship is a combination of many components, which often confuse even seasoned censorship watchers. The following is a brief overview.

- GFW (Great Firewall of China), this is a vast infrastructure deployed on the network backbone, aimed to interrupt network traffic between China and the world.
- Government-imposed content filtering in the country. China authority frequently gives out orders to media (including offline newspapers and online media including private internet companies).
- Self-censorship of domestic content providers. Often times it is hard to distinguish between self-censorship and government-imposed censorship, because the government tries very hard to hide their command trails. Also note that voluntary self-censorship is very prevalent among Chinese internet companies because many of them share the same ideology with the Party authority.
- Government-imposed content filtering for international companies. I assume Microsoft is of this case.

Conventional wisdom is that China authority exerts more censorship demand on domestic companies than international ones, because international companies are not in full control by China authority, so they may not comply as willingly, or even backfire if things go wrong. However, in this study, even to my own surprise, we found that Bing filters a big quantity of domestic content, whose existence is tolerated by the China authority and most are not filtered on Baidu. Great care has been exercised to ensure the validity of this claim and we retain all the evidences.

3 Methodology

The methodology is comprised of the following parts:

- Find a corpus of words to query the Bing search engine.
- Scraping the SERP for the terms from inside China and outside China.
- Analyze scraping results.
- More rounds of scraping to investigate and confirm hypotheses and verify blacklist entries.

3.1 Initial Corpora

In this study, we used the following sets of query terms for our investigation.

- My own research on GFW’s Wikipedia blacklist [2]. This list contains 691 terms whose Wikipedia article is blocked by GFW.
- China Digital Time’s Weibo sensitive list [7]. This list contains those terms which have undergone censorship on Weibo and it contains 1894 terms.
- Wikipedia has an article on the list of censored words in China [8]. The page contains a list for GFW terms for Google, and several lists revealed in the past, including Netease (379 terms), CCTV (1412 terms), QQ (986 terms), VeryCD (2648), and a list from the ConceptDoppler study (136). Note that there is overlap among these lists.
- Many other lists I pulled manually. For example, a list from Baidu’s hot search history archive (200 terms), and several online lists of forbidden pornography novel titles.
- During the investigation, I also handcrafted many terms to test Bing’s censorship practice.

3.2 Scraping

To determine which country edition to serve a user query, Bing relies on the user’s cookie setting. If the country preference is not present, it falls back to user IP. In order to get the China edition when scraping from outside China, we add the following line in the cookie file which ensures that we get the China serving.

```
.bing.com TRUE / FALSE 0 _FP mkt=zh-CN&ui=zh-CN&hta=on
```

There are many websites which list proxies by country. We used a large number of China proxies and non-China proxies in this study. Proxies come and go, so we need some care in our scraping program to loop through our proxy list. Besides, Bing might deny service if an IP issues too many queries, which is another issue we need to keep in mind. We counter this by limiting the querying rates and spreading out the proxies we use.

For each query term, we try to scrape the first five pages. This is done by issuing an initial request, then parse the SERP to find the appropriate URL for the next page, and so on. If no such link is found or we have reached the fifth page, we end the scraping session. All pages retrieved are saved locally for analysis. We keep cookies for these scraping sessions, to ensure a coherent search experience.

Often times, we scrape a query term multiple times, since there is natural randomness in the results received.

Also note that we run all scraping in HTTPS and with HTTPS proxy. This is to avoid GFW interference.

3.3 Analysis

Analysis is basically parsing the html files of the SERP pages. We look for three things. First, we check whether there is any “censorship notice” on the page. Second, we examine the page’s HTML source to see the underlying HTML tags and link arguments. Third, we compare the search results from inside-China and outside-China and examine the difference.

Furthermore, we run two types of scraping to identify and verify Bing’s site-level filtering and URL-level filtering. For site-level filtering, we use the `site:` operator (e.g. `site:ja.wikipedia.org`); for URL-level filtering, we use the `url:` operator (e.g. `url:en.wikipedia.org/wiki/1989`). The `site:` operator will only return

search results whose domain part matches the specified site and the `url :` operator will only return search result whose URL matches the specified url. If truly no results are found, Bing shows a no-results-found page with a missing-child information box. If there is result but the result is censored, Bing will show the censorship notice. The presence of such censorship notice serves as evidence of different types of censorship.

4 Bing Censorship Details

4.1 Types of Bing Censorship

With the wealth of scrapes for comparison, we obtained a very clear picture of Bing's censorship practice.

Overall, Bing has query-based censorship and result-based censorship. Bing's adult/safe-search filtering is query-based as well, but it has distinct triggering and presentation. In this report, unless specifically mentioned, the adult/safe-search filtering is ignored.

- For query-based censorship, the SERP page shows a censorship notice with no search results. It is clear that the query term triggers the censorship. There are three censorship notices for this case. This type only occurs for inside-China users.
- For result-based censorship, the SERP page shows some search results with a censorship notice. It means some search results are removed from the page. In the cases when all qualified search results are filtered, the page shows no results (still with the censorship notice). In this case, the censorship notice is different to the query-based censorship notice, so we can still tell that it is result-based filtering. There are four censorship notices for this case. Furthermore, result-based filtering are carried out in two fashions, site-level filtering and url-level filtering.
- Adult/safe-search filtering. This is also query-based, but it is very different to the other query-based censorship, e.g. the censorship message is very different. The adult filtering triggers only on English terms, are mostly consistent between inside-China and outside-China.

Table 4.1 below lists the censorship notice messages Bing currently displays on the SERP.

Beneath the visible censorship notices, we also examine the underlying HTML source of the page. There are three types of HTML tag for the censorship notice, `<div class="sc_error">`, `<div id="results_removed">`, and `<div id="adultblock">`. In addition, all types of censorship notices (except for QB-porn (adultblock), contain a link to Bing's online help page at <http://onlinehelp.microsoft.com/zh-CN/bing/ff808447.aspx>. More intriguingly, this href tag also contains a no-op `h=` argument like `h="ID=FD,115.1"`. This `h=` argument takes on a set of different values, let me call them *hID*. The *hID* takes a single value for most cases, except for Query-based filtering with the `<div id="results_removed">` tag for 5 terms. What does this mysterious *hID* value represent remains an open ended question.

Table 4.2 lists the possible pairs of censorship notice and the underlying HTML tags. Among them, the combination of QB and "results_removed" tag is most peculiar. Among the 149 query terms that trigger QB, only 5 of them have the "results_removed" tag. Furthermore, different scrapes for these 5 terms consistently give the "results_removed" tag, but the *hID* field change from time to time.

Upon close examination of these censorship notices for thousands of query terms, we can learn that Bing uses three types of filtering rules, plus adult/safe-search filtering.

- For query-based censorship, there is a list of terms that Bing blocks completely. The SERP is a simple censorship notice with no search results. I call them *forbidden words*. See Section 4.2 for details.
- There is a blacklist of websites whose pages are removed by Bing from its SERP. I call them *sunken sites*. See Section 4.3 for details.
- There is blacklist of URLs that are removed by Bing from its SERP, but the domain is not on Bing's site-level blacklist. I call them *vanished URLs*. See Section 4.4 for details.
- There is a list of alphabetic porn terms that Bing filters for both inside-China and outside-China users.

| Censorship Type | Short Name | Message and Translation |
|-------------------|-------------|--|
| Query-based | QB-standard | 由于中国法律法规的限制，我们删除了这些搜索词的结果。有关详细信息，请参阅 此处 。 [Translation] Because of constraints of Chinese laws and rules, we removed results for these search terms. For detailed information, please see here . |
| Query-based | QB-local | 由于本地法律法规的限制，我们删除了这些搜索词的结果。有关详细信息，请参阅 此处 。 [Translation] same as above, except changing “Chinese” to “local”. |
| Query-based | QB-rare-alt | 根据本地法律的要求，搜索结果未予显示。有关详细信息，请参阅 此处 。 [Translation] According to the requirement of local laws, search results are not shown. For detailed information, please see here . |
| Result-based | RB-standard | 依据中国的法律法规，我们屏蔽了特定的相关搜索结果。有关详细信息，请参阅 此处 。 [Translation] According to Chinese laws and rules, we blocked certain relevant search results. For detailed information, please see here . |
| Result-based | RB-local | 依据本地的法律法规，我们屏蔽了特定的相关搜索结果。有关详细信息，请参阅 此处 。 [Translation] same as above, except changing “Chinese” to “local”. |
| Result-based | RB-rare-alt | 根据本地法律的要求，部分搜索结果未予显示。有关详细信息，请参阅 此处 。 [Translation] According to the requirement of local laws, some parts of search results are not shown. For detailed information, please see here . |
| Result-based | RB-new | 某些结果已被删除 [Translation] Some results have been removed |
| Adult/Safe-search | Adult | 您所在的国家或地区要求使用严格的必应Bing 安全搜索设置，该设置将过滤掉与成人内容相关的搜索结果。若要详细了解有关 您所在国家或地区 的安全搜索要求，请参阅 必应Bing 提供搜索结果的方式 。 [Translation] The country or region you are located at requires strict Bing safe search setting, this setting will filter search results related to adult content. If you want to understand in details the safe search requirements for the country or region you are located at , please see Bing's approach of providing search results . |

Table 4.1: Bing censorship notices (blue texts denote links)

The site-level filtering and URL-level filtering might confuse people for certain cases. It helps to elaborate a bit more. The site-level blacklist consists of site names like “news.bbc.co.uk”, and the URL-level blacklist usually consists of URLs like “cn.nytimes.com/china/20130718/c18xuzhiyong/”. However, Bing’s URL-level blacklist does contain many entries like “realestateads.nytimes.com” which are site names. For these site names which are in the URL-level blacklist, but not in the site-level blacklist, it means the site will not show up in Bing SERP, but subpages on that site can still be shown (as long as that particular page is not itself in Bing’s URL-level blacklist). This sounds complicated, but they can all be determined and verified by the `site:` and `url:` scraping.

There is one caveat though. If all pages of a site indexed by Bing are censored, then we do not have a way to tell if the site itself is on Bing’s site blacklist, or all the indexed pages are on Bing’s URL blacklist. This is very rare though, since Bing’s index is quite large and usually contains more than a few pages for any site.

Another case might be that a site is filtered at the site level and at the same time certain pages on the site are filtered at the URL level as well. In this case the site-level filtering masks the URL-level filtering and we can’t make any inference on the possible URL-level filtering. If Bing removes the broad site-level filtering, then we can learn more about the URL-level filters.

One competing hypothesis is content-based filtering, which means that Bing does not use simple blacklists, but

| Censorship Notice | HTML tag & hIDs | Comment |
|-------------------|---|--|
| QB-standard | <code><div class="sc_error"></code> FD, 115.1, FD, 123.1 | most common case for QB |
| QB-standard | <code><div id="results_removed"></code> various values like SERP, ????.1 | rare |
| QB-local | <code><div id="results_removed"></code> various values like SERP, ????.1 | rare |
| QB-rare-alt | <code><div id="results_removed"></code> various values like SERP, ????.1 | rare |
| RB-standard | <code><div id="results_removed"></code> SERP, 5062, 1 | most common case inside-China |
| RB-local | <code><div id="results_removed"></code> SERP, 5062, 1 | inside-China only |
| RB-rare-alt | <code><div id="results_removed"></code> SERP, 5062, 1 | inside-China only |
| RB-new | <code><div id="results_removed"></code> SERP, 5062, 1 | outside-China only, replacing RB-standard (in-completely). |
| QB-porn | <code><div id="adultblock"></code> NA | consistent between inside-China and outside-China; only trigger on certain English terms |

Table 4.2: Examine Bing censorship notices.

instead, it studies the page content (title, snippet, all texts on the page) either real-time or periodically for everything in its index, and using the content information to decide whether to show to China users or filter. This is very unlikely. One reason is that such “smart filtering” is not an easy engineering task and may not satisfy China authority. The other reason is that we have observed lots of “sensitive” words on SERP pages scraped from inside China. Overall, the content-based filtering hypothesis is unlikely to be true.

Below we examine these different types in details.

4.2 The Forbidden Words

In this study, we identified 149 terms that are on Bing’s forbidden words list. We can not claim this list is complete, but given that we have combed through almost all major censorship lists available online, I am reasonably confident that this list covers a major chunk of Bing’s actual list. (Note, 10 terms were removed between Feb 20 and Feb 28.)

This list is “exact” match, that is, user query has to be the exact string to trigger the QB censorship notice, up to upper-case/lower-case variations or whitespaces before or after the string. If it is a two-part term (i.e. with a whitespace in the middle, there are 9 such cases), adding more whitespaces between the two parts will still trigger the QB censorship notice. On the contrary, breaking a term up by adding whitespaces, or swapping the two parts for a two-part term, or removing the whitespace from a two-part term, or combining two forbidden words into one search term, will no longer trigger QB censorship notice (unless the new composition itself is one of the forbidden words).

We can roughly group this list into three sets: political, porn, and a couple miscellaneous ones. The porn set can be further divided into two subsets: generic porn terms and titles of contemporary pornographic novels. Table B.1 in Appendix B is the complete list ordered by pinyin in each subset. Note that the porn set here is totally different to the list which triggers Bing’s adult/safe-search filtering.

As mentioned in Table 4.2, there are two HTML tags for query-based filtering, i.e. `<div class="sc_error">` and `<div id="results_removed">`. The tag is consistent with respect to query terms, i.e. querying the same term always gives the same tag. There are 5 terms in the “results_removed” category, and all the rest are in the “sc_error” category. Even though the 5 terms always give the same div tag, they often produces different hID’s (all others give a single hID). It remains a mystery why these terms are so peculiar. Table 4.3 shows these 5 terms and the observed mysterious hID’s (note, each term has been scraped ~ 10 times). Note, 3 out of these 5 are among the 10 terms which were removed between Feb 20 and Feb 28.

Query-based filtering only shows up in inside-China scraping, which is expected.

| Term | Observed hIDs | Translation & Comment |
|-------|--|---|
| 18大 | 5299.1, 5368.1, 5378.1, 5380.1, 5386.1 | 18 big, refers to CCP’s 18th National Congress (Nov 2012) |
| 六合彩* | 5402.1, 5442.1, 5447.1, 5471.1 | Mark Six, a popular lottery in Hong Kong |
| 品色堂* | 5299.1, 5304.1, 5317.1, 5334.1, 5335.1 | Sex-tasting House (an adult website) |
| 一夜情交友 | 5340.1, 5383.1, 5389.1 | find one night stand friend |
| 少妇白洁* | 5313.1, 5317.1 | Lady Bai Jie, a very popular contemporary erotic novel |

Table 4.3: Five peculiar cases. The three terms with asterisk sign were removed between Feb 20 and Feb 28.

4.3 The Sunken Sites

In this study, we identified a list of 329 “Sunken Sites” that Bing removes from its SERP.

To obtain this list, we use the following procedure.

- First, we parse all the scraped SERPs, for each website, we tabulate the numbers of occurrences in the outside-China scraping and the inside-China scraping. A pattern emerges instantly, for example, www.youtube.com showed up more than 2000 times in our outside-China scrapes but 0 time in our inside-China scrapes. We gather all websites whose outside-China counts significantly exceed their inside-China counts.
- In the next step, we run scraping for strings `site:${sitename}` (multiple times for more confidence). The `site:` prefix asks Bing to only return search results under that site. If the inside-China scrapes for `site:${sitename}` give 0 result (and with censorship notice) but the outside-China counterpart give non-zero results, we will be quite certain that this website is on Bing’s sunken sites list.
- Another source is from studying the URL-level filtering. For those URL terms whose `url:${urlstring}` querying returns 0 result (and with censorship notice), we also examine its domain part, i.e. to see what results the `site:{domain}` query returns and whether it triggers censorship notice.

There is one caveat in the `site:${sitename}` scraping. Sometimes the inside-China scrapes do give non-zero results, however, the domain is not exactly the website we are querying for, but instead it is a superstring (thus it satisfies Bing’s “site:” operator). For these cases we examined the scrapes closely. If the website is valid (e.g. the outside-China scrapes return results whose domain is exactly the website), but all inside-China scrapes give superstrings, then we can also draw the conclusion that the website is on Bing’s sunken sites list.

For example, when we query “`site:www.takungpao.com`” from inside China, we do obtain results but the results are on “`www.takungpao.com:82`”. Another example is the query “`site:cdjp.org`”, the inside-China scrapes only give results from “`public.cdjp.org`”, more investigation reveals that both “`cdjp.org`” and “`www.cdjp.org`” are on Bing’s sunken sites list but “`public.cdjp.org`” is not. Also, if a result is on “`https://`” then it will pass the site-filtering as well. An example for this case is “`www.falundafa.org`”. The “`site:www.falundafa.org`” search from inside China returns one result which is “`https://www.falundafa.org`”.

There are more than 20 such cases observed. These corner cases tell us that Bing’s site-based filtering is exact match, i.e. it removes the result if and only if the URL’s domain part matches an entry on the sunken sites list exactly. These corner cases also reinforce our understanding of Bing’s filtering mechanism.

Table C.1 in Appendix C lists all sunken sites identified in our study. They have all been verified by multiple scrapes. For those significant cases (e.g. listed below), we take even greater care to ensure they are accurate.

Most of these sites are those not enjoyed by the China authority. They include sites related to 1. news sites; 2. human rights advocacy sites; 3. overseas Chinese forum sites; 4. Falungong sites; and 5. general knowledge sites.

A few notable or odd cases are:

- Among these sunken sites, the five Wikipedia sites are the most egregious. The five versions are: Japanese, French, German, Dutch and Swedish. Fifty other versions (including English and Chinese) are not censored at the site-level (though many of their articles are censored at the URL level). In my opinion, these five sites being removed from Bing SERP entirely is a slap in the face of Microsoft to their claim “*if we are required to implement the request, we will do so narrowly*”.

- The sites www.takungpao.com and cq.takungpao.com are on the sunken sites list. Ta Kung Pao is a Hong-Kong based newspaper controlled by the CCP. It is very odd that China authority would take the effort to reach out to a foreign company to censor its own mouthpiece. The cq.takungpao.com is the site's Chongqing version. Chongqing is the city governed by the infamous Bo Xilai before his downfall in 2012. My hypothesis is that the blacklisting of www.takungpao.com might be related to the Bo Xilai case, and like many censorship restrictions, it's long forgotten and nobody knows about it any more.
- Another similar example is the site-level censorship for dailynews.sina.com, which is the international version of Sina, a major web portal in China. Sina's international version also suffers from several URL-level filtering rules. The blacklisted site/URLs identified in this study are not sensitive at all. It is another prime example of negligence from a blackbox operation with no transparency or oversight at all.

In Section 5 and Section 6, we will deep dive into two cases: New York Times and blog.163.com, a popular blogging site in China. After we see so much domestic content being censored on Bing, we will no longer be surprised by its filtering of www.takungpao.com and dailynews.sina.com!

4.4 The Vanished URLs

Besides the site-level blacklist, Bing has a very very large URL blacklist. In this study, we identified 1593 URLs blacklisted by Bing. They are listed in Table D.1 in Appendix D.

Similar to the query-level and site-level filtering, Bing's URL-level filtering is also "exact match". For example "zh.wikipedia.org/wiki/王丹"¹ is on Bing's URL blacklist, but superstring URLs like "zh.wikipedia.org/wiki/王丹凤" and "zh.wikipedia.org/wiki/王丹桂" are not filtered by Bing. Bing does ignore the trailing slash "/" and the named anchor part after "#" if there is any.

In particular, I examined most entries in GFW's Wikipedia rules and found 271 on Bing's URL blacklist. Out of these 271 entries, 223 are on zh.wikipedia.org (Chinese), 31 are on en.wikipedia.org (English), 15 are zh-yue.wikipedia.org (Cantonese), 1 on zh-classical.wikipedia.org and 1 on simple.wikipedia.org. Bing's Wikipedia rules have some overlap with GFW's Wikipedia rules, but it is not a subset. This reinforces our understanding that GFW operation is quite "remote" to China censorship bureaucracy's other branches. I have not checked all variations of Wikipedia article URLs or those terms not on GFW's list (unless it is discovered from my general scraping). So the actual count of Bing's Wikipedia rules should be much larger than reported here.

I believe what we have surfaced here is just the tip of the iceberg. I only scraped the first 5 pages for these queries, and I believe there are tons of blacklisted URLs that do not make the first 5 pages for the query terms I examined, or do not make to the SERP at all, just due to the vast body of online information. In Section 5 and Section 6 we dug deep for two cases (New York Times and blog.163.com), and surfaced many sites/URLs on Bing's sunken sites and vanished URLs lists.

4.5 Adult/Safe-Search Filter

This category is an entirely separate category. The censorship notice is distinct and the triggering is mostly consistent between inside-China and outside-China querying. It is not the focus of this study, so I did not attempt to find an extensive list. Note that it is also exact match to trigger this type of censorship. Below is a sample of query terms in this category.

| | | | | |
|-------------|------------|-------------|-----------------|-----------|
| 97sese | anal | asiasex | BlowJobs | child sex |
| fuck | hand job | hardcore | hardcore movies | hentai |
| hotsex | kids porn | naked girls | playboy | porn |
| porn videos | sex videos | tokyohot | Tokyo-Hot | xxx |

We see that no Chinese terms are in this category. Bing does perform censorship for many Chinese porn terms, but those are implemented as general query-based filtering (see Section 4.2 or Table B.1) for more details.

¹王丹: student leader in Tiananmen Square Protest in 1989

5 Deep Dive: New York Times

This section tells a slightly deeper study for New York Times (“cn.nytimes.com” and “www.nytimes.com”). Bing should have similar censorship handling for all other major news sites.

The standard procedure explained earlier revealed 18 blacklist entries for the newspaper. For the deeper study, I tried two things. First is to scrape the first 20 pages for `site:cn.nytimes.com` and `site:www.nytimes.com` from inside-China and outside-China. Second is that I query `site:cn.nytimes.com ${topic} where ${topic}` takes “新公民”², “许志永”³, “习近平”⁴, “周永康”⁵ and “奥巴马”⁶. Furthermore, I examined all the first-level and second-level subcategories for `www.nytimes.com` and `cn.nytimes.com`.

In total, we identified 33 Bing rules for New York Times. Many are news articles, some are category pages. Some are expected, e.g., the reporting on family wealth of Chinese top officials like Xi Jinping and Wen Jiabao, or Chinese social movement like Dr Xu Zhiyong and the New Citizens’ Movement. Some are quite odd though, for example, why on earth does Bing filter the Sports index page and the real estate ads page??

For what it’s worth, there is only one site-level filtering rule identified so far. It is for “aelb.cn.nytimes.com”. The site “aelb.cn.nytimes.com” looks identical to “cn.nytimes.com”, and I am unsure what is the difference between the two. All the others are URL-level filtering rules, e.g. the URL-level rule “cn.nytimes.com/” does not filter subpages under `cn.nytimes.com/`, it only filters the result “cn.nytimes.com/”.

New York Times articles have date in the URL. Glancing over the dates, it suggests that Bing deploys these filtering rules quite frequently. It could be that China authority sends multiple requests in batch mode, but my gut feeling is that Bing receives such request with very high frequency.

The identified rules are listed below.

Site-level filtering for nytimes.com

aelb.cn.nytimes.com

URL-level filtering for nytimes.com

| | |
|---|--|
| realestateads.nytimes.com | yes, real estate ads is censored |
| jobmarket.nytimes.com/pages/jobs/index.html | |
| www.nytimes.com/pages/business/index.html | |
| www.nytimes.com/pages/opinion/index.html | |
| www.nytimes.com/pages/sports/index.html | yes, the sports page is censored |
| www.nytimes.com/pages/todayspaper/index.html | |
| www.nytimes.com/pages/world/asia/index.html | |
| www.nytimes.com/pages/world/index.html | |
| www.nytimes.com/2012/10/26/business/global/family-of-wen-jiabao-holds-a-hidden-fortune-in-china-chinese-version.html?pagewanted=all | |
| cn.nytimes.com/ | |
| cn.nytimes.com/world/ | |
| cn.nytimes.com/usa/ | |
| cn.nytimes.com/policy/ | other 58 first-level subdomains are not filtered on Bing |
| cn.nytimes.com/china/20131216/c16zhou/ | related to Zhou Yongkang |
| cn.nytimes.com/china/20130902/c02corruption/ | on corruption, related to Zhou Yongkang |
| cn.nytimes.com/china/20130831/c31bo/ | related to Bo Xilai |
| cn.nytimes.com/china/20130831/c31bo/zh-hant/ | same article as above (Traditional Chinese version) |
| cn.nytimes.com/china/20130718/c18xuzhiyong/ | article on Xu Zhiyong |
| cn.nytimes.com/china/20130221/c21tibet/ | on two Tibetan self-immolation cases |
| cn.nytimes.com/china/20130215/c15tibet/ | on the 100th Tibetan self-immolation case |

²新公民: New Citizens (movement), a social movement initiated by Dr Xu Zhiyong.

³许志永: Xu Zhiyong, a legal scholar/activist who was sentenced to 4 years in prison in February 2014.

⁴习近平: Xi Jinping, CCP’s current Party head.

⁵周永康: Zhou Yongkang, CCP’s ex-security tsar under investigation right now.

⁶奥巴马: Obama, President of USA.

| | |
|--|---|
| cn.nytimes.com/world/20130215/c15xi/ | article on Xi Jinping's warning re U.S.S.R's downfall |
| cn.nytimes.com/china/20130124/c24jiang/ | related to Jiang Zemin |
| cn.nytimes.com/china/20121206/cc06moyanpress/ | on Mo Yan, the Nobel laureate |
| cn.nytimes.com/china/20121127/c27pinganside/ | related to the Wen Jiabao-Ping-An scandal |
| cn.nytimes.com/china/20121127/c27pingan-statement/ | Ping-An statement regarding Wen Jiabao |
| cn.nytimes.com/china/20121120/c20corruption/ | Xi Jinping warns about corruption |
| cn.nytimes.com/china/20121116/c16bio/ | on Xi Jinping's ascending in CCP power structure |
| cn.nytimes.com/china/20121114/c14princeling/ | on the rise of princelings in China |
| cn.nytimes.com/china/20121029/c29wen/ | on Wen Jiabao's response to NYT reporting |
| cn.nytimes.com/china/20121026/c26princeling/ | David Barboza's reporting on Wen Jiabao's family wealth |
| cn.nytimes.com/china/20121023/c23reform/ | on Xi Jinping's meeting with Hu Deping |
| cn.nytimes.com/china/20120915/c15xi/ | on Xi Jinping's reappearance after two weeks of absence |
| cn.nytimes.com/china/20120823/cc23dongyuyu/ | on Bo Xilai case |

6 Deep Dive: blog.163.com

As mentioned earlier, one striking finding in this study is that Bing filters a large amount of domestic content. The example we use in this section is blog.163.com, a popular blogging site in China, operated by a major Chinese internet company NetEase.

These sites and URLs are obtained from comparing all the blog.163.com results from outside-China scrapes and inside-China scrapes. Note that due to natural serving variation, many URLs appearing only in our outside-China scrapes are not actually filtered. Below are the items we verified through `site:` and `url:` scrapes, there are 13 site rules and 30 URL rules. Note that some of the URL rules are the site name; as explained earlier, this means that the particular URL is filtered, but subpages on that site are not.

Browsing these filtered sites and pages, we actually see very few of them are real “sensitive”. Even though some of them touch on sensitive issues, they are mostly in accordance with CCP propaganda. This is not surprising, since truly “sensitive” articles would have been “harmonized” long ago, after all, this is a domestic blogging site where China authority has full control. One noteworthy entry is the blog of Professor He Weifang, a prominent legal scholar at Beijing Univesity. His Sina blog is also on Bing's filtering list. However, neither is censored on Baidu.

Even though we tested hundreds of blog.163.com sites and pages, these are still just a tiny portion of content on blog.163.com. My gut feeling is that Bing's filtering rules for blog.163.com are an order of magnitude larger.

Site-level filtering for blog.163.com

| | |
|-------------------------------|---|
| candylulu007.blog.163.com | 颜如玉's personal blog |
| canxue911.blog.163.com | canxue911's personal blog |
| cskun1989.blog.163.com | 蔡慎坤's blog (a popular author) |
| fjq6287540.blog.163.com | 朝说图's personal blog |
| hewEIFang2009.blog.163.com | blog of He Weifang, a prominent legal scholar |
| lxc66188.blog.163.com | 理睬's personal blog |
| npfj.blog.163.com | 痴心小妹's personal blog |
| oushi1937.blog.163.com | oushi1937's personal blog (pro-Maoism) |
| sebowuyuetianxin.blog.163.com | 色播五月天's blog |
| vip999sw.blog.163.com | 吝色鬼's blog |
| whh631211.blog.163.com | 晋商1211's personal blog (pro-Maoism) |
| wmy1923.blog.163.com | 展望未来's personal blog (unaccessible) |
| yuzhengsheng2009.blog.163.com | a personal blog using the Party leader Yu Zhengsheng's name 俞正聲 |

URL-level filtering for blog.163.com

| | |
|--|-------------------------------|
| 19210701zggcd.blog.163.com/blog/static/306696392008111202812837/ | article denouncing Charter 08 |
|--|-------------------------------|

| | |
|---|------------------------------------|
| allenhe.ok.blog.163.com/blog/static/2534573420106178658978/ | old article on UltraSurf |
| bairuohong2007.blog.163.com/blog/static/520785222012012661736/ | a hidden article |
| chensuleon.blog.163.com/blog/static/1409735042010111795652278/ | an English email on Xi Jinping |
| cykctadcl.blog.163.com/ | 哥们干杯的博客's personal blog |
| dinglixin971.blog.163.com/ | 快乐是一种心情's personal blog |
| ftmouse.blog.163.com/blog/static/261041201010283533753/ | article on “朝” |
| hechangbao123.blog.163.com/blog/static/2363217920101224513129/ | a joke |
| hongzhi528.blog.163.com/ | hongzhi's blog, on computer stuff |
| jibeixiongguan.blog.163.com/blog/static/2772833020093177552913/ | a joke involving Song Zuying |
| jinduba.blog.163.com/ | 品色堂's personal blog |
| ksd08688689.blog.163.com/ | 宁波老蒋的博客's personal blog |
| leebapa.blog.163.com/blog/static/11096813820095992554259/ | a negative news on Wuer Kaixi |
| lhbb8.blog.163.com/blog/static/64745841200951952740110/ | article on Yu Zhengsheng |
| lihongtao90.blog.163.com/blog/static/97624946200911715624790/ | article denouncing student leaders |
| lipeng.868.blog.163.com/ | 月月鸟's blog, unrelated to Li Peng |
| lixiaocheng666888.blog.163.com/blog/static/9717075720101117105720456/ | article denouncing Xin Ziling |
| lixiaocheng666888.blog.163.com/blog/static/97170757201032253442138/ | article denouncing Zhao Ziyang |
| rainysia.blog.163.com/blog/static/14191021020101123105710585/ | article on Google and IPv6 |
| shhuanyayq.blog.163.com/blog/static/9754761420099771623710/ | article on TOR |
| sunny-9196271.blog.163.com/ | blog of 牛刀, a popular blogger |
| sxszl.blog.163.com/blog/static/54494648200702844725393/ | plain biography of Bao Tong |
| wangwanxin0814.blog.163.com/blog/static/1015187932008101610655328/ | article on Yu Zhengsheng family |
| wenroudeyu927.blog.163.com/ | 温柔细雨's personal blog |
| wzhsh888.blog.163.com/blog/static/134347208201235115654147/ | article criticizing Jiang Zemin |
| x1957613.blog.163.com/blog/static/168090036201341582615474/ | page inaccessible |
| xysalon.blog.163.com/blog/static/5235292007542194207/ | analysis on June 4th (Party line) |
| yingyingylf.blog.163.com/blog/static/75666020084323948550/ | article on Hu Jintao's PKU visit |
| yinjian.688.blog.163.com/blog/static/8450985920104184916415/ | page inaccessible |
| zhy-xianger.blog.163.com/ | 香儿's personal blog |

In addition to blog.163.com, we tested other major domestic sites. Lo and behold, all of them have something censored by Bing! Table 6.2 summarizes what our study has identified, which is only the tip of the iceberg. For the actual filtering rules, see Table D.1 in Appendix D, or the “Vanished URLs” tab on the master spreadsheet <http://goo.gl/qzAaVe> [1].

7 The “Leakage” and Gray Censorship

7.1 The “Leakage”: Bing's censorship notice outside China

In this section, we discuss the much-debated accusation that Bing performs China censorship outside China. This is a very serious accusation and Microsoft denied vehemently, while at the same time GreatFire.org researchers showed solid evidence supporting their claims. My study does not answer this question definitely, but it sheds some light on how prevalent this problem is.

GreatFire's most strong evidence is that freeweibo.com's homepage is not shown up when searching for “freeweibo.com” outside China. Together with Bing's official response, it is clear that “freeweibo.com” had once been put on Bing's global URL blacklist. Bing attributed this to an “error” that it mistakened the freeweibo.com page as a spam website.

Another piece of strong evidence by GreatFire is that Bing displayed the specific China censorship notice for certain searches performed outside China. Bing's explanation is that the removals were not due to China censorship, but instead, the removals were due to pornography, copyright, or other issues, and Bing simply messed up with the censorship messaging, i.e., it mistakenly used the China censorship notice for these other types of censorship.

| Site | Rule Count | Description |
|------------------|------------|--|
| blog.163.com | 43 | NetEase blog |
| 360doc.com | 19 | a file storage and sharing service |
| 56.com | 2 | a video hosting site |
| baike.baidu.com | 31 | Baidu baike (a Wikipedia knock-off) |
| hi.baidu.com | 2 | Baidu blog |
| tieba.baidu.com | 5 | Baidu discussion forum |
| wenku.baidu.com | 5 | Baidu library |
| zhidao.baidu.com | 7 | Baidu question/answer service |
| douban.com | 29 | Douban, a review site |
| ifeng.com | 17 | a major portal site |
| qq.com | 19 | a major portal/personal blogging site |
| blog.sina.com.cn | 26 | Sina blog, China’s biggest blogging site |
| baike.soso.com | 10 | Soso Baike (a Wikipedia knock-off) |
| bbs.tianya.cn | 5 | Tianya bbs, a major forum site |
| tudou.com | 4 | Tudou, a video hosting site |
| youku.com | 12 | Youku, a video hosting site |

Table 6.2: Bing’s censorship on domestic content. Counts are only of the small set we identified in this study. Counts include both site-level rules and URL-level rules.

After the debacle, Bing’s fix is to replace the China censorship notice by a generic censorship notice, i.e. changing from RB-standard to RB-new (see Table 4.1) to users outside China.

Unlike my study on Bing’s censorship in China, it is much harder to figure out what have been removed from Bing globally. Nevertheless, in my study, we have gathered tens of thousands of SERPs from querying Bing from outside China. Among the query terms we studied, Bing has shown censorship notice for 1,710(!) unique query terms (within the first five pages of SERP) to users outside China. The censorship notice are RB-new (majority) and RB-standard (minority). This count does not include those English porn terms which trigger the QB-porn message.

These 1,710 terms cover a very wide range of topics. Roughly one third of them are porn, and the rest are political. Appendix E lists 644 items out of the political subset (we removed some long terms and terms with traditional Chinese characters to make the table manageable). For curious readers outside China, one can set Bing’s version to Chinese, then search any of these terms to look for the censorship notice at the bottom of the page. Note that the censorship notice may not show up on the first page (we scraped up to 5 pages). Three examples I use often are “jiang zemin”, “达赖喇嘛” and “Dalai Lama”, for which the censorship notice appears on the first page.

It is near-impossible for an outsider to figure out what is going on. For these many political terms which trigger censorship (outside China), what are the filtered results? Since the censorship notice shows up within the first five pages (and often on the very first page), it means at least Bing thinks these results are highly relevant, otherwise they won’t get on the first page or the first five pages. The reasonable suspicion is that these are not all porn results, spammy website, or copyright-violating websites. I think Microsoft has an obligation to explain what is happening.

7.2 Gray Censorship

So far, we have been focusing on SERPs with censorship notice shown. Bing has been faithful in displaying these notices when there are results being filtered, for this I think we should commend Microsoft.

However, the “gray censorship” might be a more serious issue. As Greatfire.org has pointed out, searching for query terms like “达赖喇嘛” (Dalai Lama) in the Chinese version from outside China gives mostly pro-CCP results. Rebecca MacKinnon talked about this issue in her Guardian article[6] in great length. It is unquestionable that Bing’s algorithm is favoring Chinese domestic sites, which are under tight control for topics like Dalai Lama. A casual reader who uses Bing’s Chinese version, will more likely to become to believe in CCP’s side of the story.

Again, as an outsider, I have no knowledge about Bing’s ranking algorithm. It could be that Bing engineers have been absolutely impartial regarding ranking about political sensitive subjects, but we can not rule out possibility that the ranking algorithm is tweaked, intentionally or unintentionally, to favor domestic websites more than they deserve. The engineers have the capability to tweak the system algorithmically to mask all censorship notices without being

caught by outsiders, and that is the “ultimate censorship” - censorship without people realizing they are being fed with censored information. Here, we ask Bing engineers to hold up their promises to support the freedom of access to information.

8 Conclusion Remarks

In this study, we thoroughly examined Bing's censorship practice for China.

It is well known that Dr Lu Qi, executive vice president of Microsoft's Applications and Services Group, which includes the Bing search engine, is from China. It is saddening that a Chinese with big responsibility in the freer world, tolerates this kind of unethical cooperation with China censorship.

I hope the revelation in this report could serve as a reminder to international corporations that cooperation with unethical censorship comes with a risk. Furthermore, treating Chinese users as not deserving the same standard of information freedom, is not just insulting, but also a crime to humanity.

8.1 My Requests to Microsoft

In the spirit of transparency and honoring Microsoft's own published stance regarding censorship (see Appendix A), I make the following requests to Microsoft:

- Explain why Bing censors so much content, including domestic content which are not even censored by Chinese companies.
- Audit the actual procedure and handling of censorship in China, to verify that all government requests are scrutinized by proper personnel. In particular, investigate the approval process for the domestic content.
- Present the proof that removing entire sites off Bing in China is based on applicable law. In particular, the removal of the entire Japanese, French, German, Dutch and Swedish Wikipedia sites.
- Do not attempt to hide censorship details, e.g. by altering censorship notices, altering how search operators like `site:` and `url:` work, or tweaking ranking algorithms.

8.2 Acknowledgements

This research is a solo project over a very short period of time, so this is a short list. The author thanks Greatfire.org for their initial revelation, internet freedom activists like Jason Q. Ng and Rebecca McKinnon for their well thought insights and opinions, and media outlets like the Guardian, WSJ, Verge etc for their reporting of the scandal.

References

- [1] Xia Chu. An Audit on Bing's China Censorship, or, an Independent Transparency Report. <http://goo.gl/qzAaVe>.
- [2] Xia Chu. The Complete List of GFW Rules for Wikipedia (with active updates), websites, IPs, IMDB, AppStore and more. goo.gl/zKslcu.
- [3] Greatfire.org. Bing在全球实行中国式审查(Bing Performs China Censorship Around the Globe) (Feb 11, 2014) zh.greatfire.org/blog/2014/feb/bing-practicing-chinese-censorship-globally
- [4] Bing official statement by Stefan Weitz. Setting the Record Straight (Feb 12, 2014) www.bing.com/blogs/site_blogs/b/search/archive/2014/02/12/update.aspx
- [5] Jason Q. Ng. Reputation Matters: Unpacking the Microsoft China Censorship Scandal. Wall Street Journal (Feb 14 2014) blogs.wsj.com/chinarealtime/2014/02/14/reputation-matters-unpacking-the-microsoft-china-censorship-scandal
- [6] Rebecca MacKinnon. Where is Microsoft Bing's transparency report? The Guardian (Feb 14, 2014) <http://www.theguardian.com/commentisfree/2014/feb/14/microsoft-bing-china-censorship-transparency>
- [7] China Digital Times. Weibo Sensitive Terms. chinadigitaltimes.net/chinese/category/网情透视/敏感词库/?view=all
- [8] Wikipedia. List of Censored Words in People's Republic of China. zh.wikipedia.org/wiki/中華人民共和國審查辭彙列表

Appendices

A Bing's Official Claims Regarding Censorship Practice

For the record, we paste Bing's official claims regarding its censorship practice below. Certain words are highlighted by the author.

The Chinese version is located at <http://onlinehelp.microsoft.com/zh-CN/bing/ff808447.aspx>.

我们如何解决各个国家/地区的具体法律

一些国家拥有适用于搜索服务提供商的法律或法规，要求我们对必应Bing 已列入索引的某些信息删除相关访问权限，这主要是出于地缘政治的目的或当地文化规范和敏感性的要求。我们必须既要支持所有国家/地区的人们自由访问信息，又要符合允许我们在特定管辖区内提供搜索服务的要求。如果政府机构要求删除显示的搜索结果与我们联系，我们需要政府机构提供适用法律和权利的证明，以及官方的删除要求申请。如果提供了这类证明且我们可以查证，那么我们可以履行删除申请。如果要求我们必须实施申请，我们将有限地实施申请。如果删除申请不符合国际标准，我们可能选择询问相关解释以进一步减少我们须履行的义务。

The English version is located at <http://onlinehelp.microsoft.com/en-US/bing/ff808447.aspx>.

How we address laws specific to individual countries

Some countries maintain laws or regulations that apply to search service providers that require that we remove access to certain information that Bing has indexed, primarily for geopolitical purposes or local cultural norms and sensibilities. We must integrate our support for freedom of access to information by people of all countries with required compliance that allows us to offer the search services in a specific jurisdiction. When approached with a request for removal of displayed search results by a governmental entity, we require proof of the applicable law and authority of the government agency, and an official request to require removal. If such proof is provided and we can verify it, then we may comply with the removal request. If we are required to implement the request, we will do so narrowly. If the removal request is inconsistent with international standards, we might choose to seek clarification to further narrow our obligation to comply.

B Forbidden Words on Bing

Below we list the forbidden words we identified in this study. Forbidden words are those terms that Bing returns an empty page with particular censorship notice. The terms marked with the pound sign are those 5 peculiar terms (explained in Section 4.2) which remain a big mystery. There are 10 terms which were removed by Microsoft between Feb 20 and Feb 28, they are marked with the asterisk sign.

| Term | Translation and Comment |
|------------------|--|
| Political | |
| 18大 # | 18 big, referring to CCP's 18th National Congress (Nov 2012). |
| free tibet | a taboo phrase in China and among many Chinese people. |
| 半羽 | half weather, referring to Xi Jinping. |
| 赤匪 | red bandit, referring to CCP. |
| 裆中央 | center of crotch. “裆” (crotch) has the same pronunciation as “党” (Party). |
| 党魁 | Party head. |
| 负面信息删除 | negative information removal. |
| 蛤蟆 病危 | toad + critical medical condition. Toad refers to Jiang Zemin. |
| 共惨党 | Party of Common Suffering. A homophonic play on “共产党” (Communist Party). |
| 共匪 | communist bandit, referring to CCP. |
| 古月金帛三寿 | a character-split play on “胡锦涛” (Hu Jintao). |
| 广隶 | a split-character play on “康” (Kang), referring to Zhou Yongkang. |
| 胡core | Hu core, referring to Hu Jintao. |
| 胡jt | referring to Hu Jintao (surname plus initials in pinyin). |
| 黄海刺胡 | assassination of Hu in Yellow Sea. A rumored event in May 2006. |
| 假庆淋 | a homophonic play on “贾庆林” (Jia Qinglin). |
| 江core | Jiang core, referring to Jiang Zemin. |
| 江zm | referring to Jiang Zemin (surname plus initials in pinyin). |
| 江核心 | Jiang core, referring to Jiang Zemin. |
| 江家帮 | Gang of Jiang. |
| 江绵恒 | Jiang Mianheng, son of Jiang Zemin. |
| 江派 | Faction of Jiang. |
| 江戏子 | Jiang actor, referring to Jiang Zemin. |
| 江贼民 | a homophonic play on “江泽民” (Jiang Zemin). The middle character means “thief”. |
| 九常委 | nine standing committee members (of CCP's politburo). |
| 九长老 | nine presbyters, referring to the nine standing committee members. |
| 九哥 | nine brothers, or brother #9. Unclear what this term means. |
| 康师傅 下架 | Master Kang + off shelf. Master Kang is a brand of instant noodle, referring to Zhou Yongkang. |
| 令jh | referring to Ling Jihua (surname plus initials in pinyin). |
| 前核心 | former core, referring to Jiang Zemin. |
| 删除负面信息 | remove negative information. |
| 上海帮 | Gang of Shanghai (a.k.a. Faction of Jiang). |
| 太子党 | party of princelings, a loose faction in CCP by descendents of previous generation of CCP heads. |
| 天线宝宝 康师傅 | Teletubbies + Master Kang. Teletubbies refer to Wen Jiabao. |
| 团派 | Faction of CCP Youth League. |
| 温jb | referring to Wen Jiabao (surname plus initials in pinyin). |
| 温秀秀 | Wen Show-Show, referring to Wen Jiabao, who is regarded by many as pretending. |
| 温影帝 | Wen best actor, referring to Wen Jiabao. |
| 瘟家宝 | a homophonic play on “温家宝” (Wen Jiabao). The first character means “plague”. |
| 五不搞 | Five Donot's. CCP's five principles against modern political reform. |
| 误邦国 | a homophonic play on “吴邦国” (Wu Bangguo). The first character means “error”. |
| 习jp | referring to Xi Jinping (surname plus initials in pinyin). |
| 习 太子 | Xi + crown prince, referring to Xi Jinping after he was appointed the successor to Hu Jintao. |

continued on next page...

...continued from previous page

| Term | Translation and Comment |
|--|---|
| 一党独裁 | one-party dictatorship. |
| 一党专政 | one-party authoritarian regime. |
| 俞正声 哥哥 | Yu Zhengsheng + elder brother, referring to Yu Qiangsheng, an intelligence official defected to the US. |
| 俞正声 家族 | Yu Zhengsheng + family. |
| 张德江 动车 | Zhang Dejiang + bullet train. Zhang ordered to bury wrecked trains in the 2011 Wenzhou train accident. |
| 中共暴政 | CCP tyranny. |
| 中国茉莉花革命 | China jasmine revolution. |
| 中国 茉莉花革命 | China + jasmine revolution. |
| 自由门 * | Freegate, a popular GFW-circumvention software. |
| Misc | |
| 出售枪支 | firearms for sale. |
| 出售手枪 | pistol for sale. |
| 六合彩 # * | Mark Six, a popular lottery in Hong Kong. |
| 炸弹制作 | bomb-making. |
| Porn (excluding contemporary novel titles) | |
| 18禁 | banned for minors under 18. |
| av网址 | av (adult video) website. |
| 北京xx网 | Beijing xx website. |
| 逼痒 | vulva itch. |
| 成人聊天室 | adult chat room. |
| 处女 * | virgin. |
| 灯草和尚 | Monk Tang Cho (a classic erotic novel). |
| 丁香成人社区 | Clove adult community (an online community). |
| 多p杂交 | multi-player sex activity |
| 肥逼 | fat vulva. |
| 国产av | domestic av (adult video). |
| 禁春 | Gam Cheun (a 1993 softcore movie). |
| 狼友 * | patron to sexual services. |
| 迷奸 | to daze and rape. |
| 品色堂 # * | Sex-tasting House (an adult website). |
| 色情聊天室 | sexual chat room. |
| 色情文学 | erotic literature. |
| 兽性新人类 | Beastly New Human Kind (a Hong Kong softcore horror movie). |
| 兽欲 | bestial desire. |
| 素人电影 | amateur (adult) movie. |
| 尾行 * | Stalker (an adult computer game). |
| 一本道 * | IPondo (a Japanese adult product company and brand). |
| 一夜情交友 # | find one night stand friend. |
| 幼逼 | young vulva. |
| 玉蒲团 * | Sex and Zen (a movie adapted from a classic erotic novel). |
| 真人做爱视频 | real intercourse video. |
| 足本玉蒲团 | Full Version Sex and Zen (another adaption of the novel Sex and Zen). |
| Porn: Contemporary Novel Titles (translation omitted) | |
| 阿兵哥言语录 | 阿櫻的悄悄话 |
| 爸爸的新娘 | 不良少女日记 |
| 沧澜曲 | 创世之子猎艳之旅 |
| 摧花神龙教 | 大丑风流记 |
| 第一无赖 | 癫马女郎之一夜情 |
| 东北风情熟女之惑 | 风骚侍女 |
| 风月大陆 | 夫妇乐园 |
| 还珠格格成人版 | 海盗的悠闲生活 |
| 黑星女侠 | 红楼遗秘 |

continued on next page...

...continued from previous page

| Term | Translation and Comment |
|-------------|-------------------------|
| 花心天子 | 幻想玉女 |
| 妓女日记 | 江湖艳情录 |
| 交换姐妹花 | 娇宠水芙蓉 |
| 娇艳江湖 | 金鳞岂是池中物 |
| 金瓶双艳 | 金瓶艳史 |
| 绝色家族 | 冷香谷 |
| 倩影蝎心 | 猎艳江湖梦 |
| 玲珑孽怨 | 骆冰淫传 |
| 美少妇的哀羞 | 迷离的嫂嫂 |
| 女学生的耻蜜 | 秦青的幸福生活 |
| 青楼十二房 | 少妇白洁 # * |
| 少年阿宾 * | 射鸡英雄传 |
| 神雕外传之郭襄 | 生人勿近之邪花 |
| 十景缎 | 十年孤剑沧海盟 |
| 双面媚娘 | 睡着的武神 |
| 四海龙女 | 天降神龙 |
| 田螺艳鬼 | 偷食女学生 |
| 偷试隔墙花 | 王子淫传 |
| 我的性启蒙老师 | 午夜女郎 |
| 现代广州白领的淫荡生活 | 现代艳帝传奇 |
| 星光伴我淫 | 艳遇编年史 |
| 倚天屠龙别记 | 淫术炼金士 |
| 淫欲城堡 | 玉麟传奇 |
| 云雨欲魔欲魔艳史 | 照日天劫 |

Table B.1: Bing's Forbidden Words List

C Sunken Sites on Bing

Below we list the sunken sites we identified in this study. Sunken sites are those sites whose pages will never be shown on Bing's SERP inside China. When such a result should have been shown on the SERP if not for the censorship, a particular censorship notice will be displayed. Section 4.2 contains the technical details.

Sunken Site

| | | |
|-------------------------------------|-------------------------------|-----------------------------|
| www.131001.com | candylulu007.blog.163.com | canxue911.blog.163.com |
| cskun1989.blog.163.com | fjq6287540.blog.163.com | heweifang2009.blog.163.com |
| lxc66188.blog.163.com | npfj.blog.163.com | oushi1937.blog.163.com |
| sebowuyuetianxin.blog.163.com | vip999sw.blog.163.com | whh631211.blog.163.com |
| wmy1923.blog.163.com | yuzhengsheng2009.blog.163.com | www.171069.com |
| www.1eew.com | www.2008bocai.com | www.4020.cn |
| bbs.51.ca | www.61172.com | 64memo.com |
| www.64memo.com | www.64memo.org | www.64tianwang.com |
| bbs.6park.com | ent.6park.com | web.6park.com |
| 72.52.81.28 | 89-64.org | www.94007.com |
| bbx.ablwang.com | www.aboluowang.com | aid-china.org |
| www.asianews.it | www.atgfw.org | www.backchina.com |
| www.bannedbook.net | www.bannedbook.org | www.beijing1989.com |
| beijingspring.com | www.beyondfirewall.com | www.bignews.org |
| bjzc.org | 08charterbbs.blogspot.com | 1989bbs.blogspot.com |
| alrider.blogspot.com | bannedbook.blogspot.com | cdmrc.blogspot.com |
| chinajasminerevolution.blogspot.com | chinanewsjournal.blogspot.com | cnjasmine.blogspot.com |
| cnlaf.blogspot.com | huanying12.blogspot.com | jasmine-action.blogspot.com |
| lihlii.blogspot.com | molihuageming.blogspot.com | molihuagm.blogspot.com |
| molihuazh.blogspot.com | panchinese.blogspot.com | program-think.blogspot.com |
| rightscampaign.blogspot.com | smgInc.blogspot.com | tadegou.blogspot.com |
| uvgov.blogspot.com | xuzhiyong2012.blogspot.com | zhongguojinwen.blogspot.com |
| zhongguomolihua.blogspot.com | www.bokerb.com | blog.boxun.com |
| boxun.com | news.boxun.com | www.boxun.com |
| www.boxunblog.com | toronto.canadameet.com | canyu.org |
| www.canyu.org | www.cccpism.com | cdjp.org |
| www.cdjp.org | cdp1989.org | www.cdp1998.org |
| cdp2006.org | www.cdp2006.org | www.cdpwu.org |
| www.chenpokong.com | www.china-week.com | tw18.blog.china.com |
| www.chinacitynews.be | chinadigitaltimes.net | www.chinaelections.org |
| www.chinagfw.org | www.chinainperspective.com | www.chinatimes.com |
| chineseopen.org | www.chineseopen.org | discuz.club1069.com |
| www.cnd.org | news.bbc.co.uk | 1688.com.au |
| sina.gc2.com.ar | www.1688.com.au | www.books.com.tw |
| www.kanzhongguo.com.au | www.open.com.hk | bbs.comefromchina.com |
| www.comefromchina.com | www.conceptdoppler.org | www.cqcounter.com |
| bbs.creaders.net | ent.creaders.net | news.creaders.net |
| www.creaders.net | dafagood.com | dafahao.com |
| www.dailymotion.com | tuidang.dajiyuan.com | www.dajiyuan.eu |
| dalailama.com | ddhw.cn | blog.donews.com |
| dongtaiwang.com | us.dongtaiwang.com | www.dongtaiwang.com |
| www.duping.net | 18.dnnews.com | blog.dnnews.com |
| china.dnnews.com | entertainments.dnnews.com | forum.dnnews.com |
| global.dnnews.com | history.dnnews.com | hongkongmacao.dnnews.com |
| life.dnnews.com | military.dnnews.com | opinion.dnnews.com |

continued on next page...

...continued from previous page

Sunken Site

| | | |
|----------------------------------|---|----------------------------|
| politics.dwnnews.com | taiwan.dwnnews.com | www.dwnnews.com |
| forum.edoors.com | news.edoors.com | www.edoors.com |
| egongguan.net | www.epicbook.com | 64poem.blog.epochtimes.com |
| au.epochtimes.com | cn.epochtimes.com | epaper.epochtimes.com |
| hk.epochtimes.com | luntan.epochtimes.com | mag.epochtimes.com |
| qzxy.blog.epochtimes.com | tuidang.epochtimes.com | tw.epochtimes.com |
| weiquanshiji.blog.epochtimes.com | www.ccpsecretorigin.blog.epochtimes.com | www.epochtimes.com |
| www.epochtimes.jp | bbs.eulam.com | bbs.faloo.com |
| big5.falundafa.org | falundafa.org | www.falundafa.org |
| www.faluninfo.net | tw.fgmtv.org | www.asp.fgmtv.org |
| yuming.flnet.org | www.freechina.net | www.freeinews.com |
| freenet-china.org | www.freenewscn.com | freetibet2008.org |
| www.freetibet.org | www.fromchinatousa.net | bbs.futurechinaforum.org |
| www.fxingw.com | globalrescue.net | godblesschina2008.org |
| www.hao122.cn | news.have8.com | novel.hongxiu.com |
| hongzhi.li | biweekly.hrichina.org | www.hrichina.org |
| www.hrw.org | www.htkou.net | www.huaglad.com |
| huanghuagang.org | www.huanghuagang.org | forum.iask.ca |
| iask.ca | www.iask.ca | ieasy5.com |
| 9jqbqsvg.blog.ifeng.com | www.inkui.com | forums.internetfreedom.org |
| www.ip134.com | chinese.ilib.ir | www.jszsf.org |
| www.neng.jumpingcrab.com | www.kugo.cc | kxz30.com |
| www.lalulalu.com | observechina.laogai.org | m.letscorp.net |
| www.letscorp.net | lishankj.com | www.malaysiakini.com |
| www.meinvktv.com | www.metacafe.com | hong.vic.mh4u.org |
| shanyuan.vic.mh4u.org | www.mhbbs.net | mhradio.org |
| www.mhradio.org | www.minghui-school.org | big5.minghui.org |
| li-hongzhi.minghui.org | library.minghui.org | m.minghui.org |
| package.minghui.org | photo.minghui.org | weekend.minghui.org |
| www.minghui.jp | www.minghui.org | www.mingjingnews.com |
| www.mingpaonews.com | minzhuzhanxian.com | minzhuzhongguo.org |
| www.minzhuzhongguo.org | city.mirrorbooks.com | www.mirrorbooks.com |
| www.mitbbs.com | beta.molihua.org | www.molihua.org |
| www.msguancha.com | news.hk.msn.com | c36c3c32.mysinablog.com |
| news.nanyangpost.com | www.naol.ca | www.nzchinese.net.nz |
| forum.netix.com | m.ntdtv.com | www.ntdtv.com |
| aelb.cn.nytimes.com | bbs.omnitalk.org | omnitalk.com |
| www.omnitalk.com | www.onlinecha.com | www.onmoon.com |
| www.onmoon.net | my.opera.com | dzedu.org.cn |
| www.alliance.org.hk | www.dzedu.org.cn | www.falundafa.org.my |
| www.mingjing.org.cn | www.tibet.org.tw | yydg.paowang.net |
| peacehall.com | www.peacehall.com | www.penchinese.com |
| www.pinpai666.com | www.popyard.com | bbs.powerapple.com |
| news.powerapple.com | 2010.qq.com | yuming.qxbbs.org |
| renminbao.com | www.renminbao.com | www2.renminbao.com |
| www3.renminbao.com | www.renyurenquan.org | www.rfa.org |
| www.chinese.rfi.fr | blog.roodo.com | www.rxhj.net |
| www.secretchina.com | shanchutieba.com | dailynews.sina.com |
| big5.soundofhope.org | soundofhope.org | www.soundofhope.org |
| studentsforafreetibet.org | cq.takungpao.com | www.takungpao.com |
| www.shop0.taobao.com | taolawnet.net | www.tiananmenuniv.net |
| www.www1.tianyablog.com | tmc-hk.org | www.toppk.net |

continued on next page...

...continued from previous page

Sunken Site

| | | |
|-------------------------|--------------------------|-----------------------------|
| www.tsquare.tv | www.ttmil.com | www.tuidang.org |
| cnjasmine.tumblr.com | www.uwants.com | www.vot.org |
| www.wangruowang.org | bbs.we54.com | news.we54.com |
| wengewang.org | blog.wenxuecity.com | www.wenxuecity.com |
| de.wikipedia.org | fr.wikipedia.org | ja.wikipedia.org |
| nl.wikipedia.org | sv.wikipedia.org | bbs.wolfax.com |
| molineews.wordpress.com | oliuxiaobo.wordpress.com | saveliuxiaobo.wordpress.com |
| xianql.wordpress.com | wujieliulan.com | www.xinsheng.net |
| xinshijue.org | www.xmeise.com | www.xys.org |
| blog.yam.com | www.yangjianli.com | www.youpai.org |
| www.youtube.com | www.yuanming.net | www.zgdsw.com |
| www.zhengjian.org | www.zhengwunet.org | www.zhuichaguoji.org |
| www.zonghexinwen.net | zonghexinwen.com | |

Table C.1: Bing's Sunken Sites List

D Vanished URLs on Bing

Below we list the vanished URLs we identified in this study. Vanished URLs are those web pages that Bing will never display in China. When such a result should have been shown on the SERP if not for the censorship, a particular censorship notice will be displayed. For URLs whose domain is censored at the site level, they are not listed here.

The vanished URL list is too long to fit nicely into this report, so we put the URL as href target, and group by domain. The actual URLs can be found in the “Vanished URLs” tab on the master spreadsheet <http://goo.gl/qzAaVe> [1]. This is only a partial list. I believe Bing's actual URL blacklist is an order of magnitude larger. Section 4.4 contains the technical details.

Vanished URLs

| | |
|---------------------------------|---|
| 19210701zggcd.blog.163.com/ | 1 |
| allenhe.ok.blog.163.com/ | 1 |
| bairuohong2007.blog.163.com/ | 1 |
| chensuleon.blog.163.com/ | 1 |
| cykctadcl.blog.163.com/ | 1 |
| dinglixin971.blog.163.com/ | 1 |
| ftmouse.blog.163.com/ | 1 |
| hechangbao123.blog.163.com/ | 1 |
| hongzhi528.blog.163.com/ | 1 |
| jibeixiongguan.blog.163.com/ | 1 |
| jinduba.blog.163.com/ | 1 |
| ksd08688689.blog.163.com/ | 1 |
| leebapa.blog.163.com/ | 1 |
| lhhb8.blog.163.com/ | 1 |
| lihongtao90.blog.163.com/ | 1 |
| lipeng.868.blog.163.com/ | 1 |
| lixiaocheng666888.blog.163.com/ | 1 2 |
| rainysia.blog.163.com/ | 1 |
| shhuanyayq.blog.163.com/ | 1 |
| sunny-9196271.blog.163.com/ | 1 |
| sxszl.blog.163.com/ | 1 |
| wangwanxin0814.blog.163.com/ | 1 |
| wenroudeyu927.blog.163.com/ | 1 |
| wzhsh888.blog.163.com/ | 1 |
| x1957613.blog.163.com/ | 1 |
| xysalon.blog.163.com/ | 1 |
| yingyingylf.blog.163.com/ | 1 |
| yinjian.688.blog.163.com/ | 1 |
| zhy-xianger.blog.163.com/ | 1 |
| www.360doc.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 |
| www.56.com/ | 1 2 |
| www.64wiki.com/ | 1 2 3 4 5 |
| www.89-64.org/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 |
| tw.aboluowang.com/ | 1 2 3 4 5 6 7 |
| www.alexa.com/ | 1 2 3 4 5 6 7 |
| www.amazon.cn/ | 1 |
| baike.baidu.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 |
| hi.baidu.com/ | 1 2 |
| tieba.baidu.com/ | 1 2 3 4 5 |
| wenku.baidu.com/ | 1 2 3 4 5 |
| zhidao.baidu.com/ | 1 2 3 4 5 6 7 |
| britemimg.blogspot.com/ | 1 2 3 4 |

continued on next page...

...continued from previous page

Vanished URLs

| | |
|--------------------------------------|--|
| chinaajsm.blogspot.com/ | 1 2 3 4 5 6 7 8 9 10 |
| duxuezhong.blogspot.com/ | 1 2 3 4 5 6 7 |
| fanqiangblog.blogspot.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 |
| janadabc.blogspot.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 |
| jasmineplaces.blogspot.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 |
| newsabeta.blogspot.com/ | 1 2 3 4 5 6 |
| wqw2010.blogspot.com/ | 1 2 3 4 |
| yyyyyiii.blogspot.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 |
| zgldzm.blogspot.com/ | 1 2 3 4 5 6 |
| zhonggongsi.blogspot.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 |
| www.botanwang.com/ | 1 2 3 4 5 |
| briian.com/ | 1 2 |
| www.cdpsite.org/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 |
| www.cenews.eu/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 |
| book.ceqq.com/ | 1 2 |
| ceqq.com/ | 1 2 |
| www.ceqq.com/ | 1 |
| chinainperspective.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 |
| www.cnn.com/ | 1 |
| www.bbc.co.uk/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 |
| blog.sina.com.cn/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 |
| blog.sina.com.tw/ | 1 |
| control.blog.sina.com.cn/ | 1 |
| wenda.google.com.hk/ | 1 2 3 4 5 6 7 8 9 |
| www.libertytimes.com.tw/ | 1 |
| www.zaobao.com.sg/ | 1 2 3 4 5 |
| blog.creaders.net/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 |
| www.dalailama.com/ | 1 |
| www.dm5.com/ | 1 |
| book.douban.com/ | 1 2 3 4 5 6 7 |
| movie.douban.com/ | 1 2 3 4 5 6 7 8 9 10 |
| music.douban.com/ | 1 2 3 |
| site.douban.com/ | 1 |
| www.douban.com/ | 1 2 3 4 5 6 7 8 |
| www.dw.de/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 |
| news.eastday.com/ | 1 2 3 4 |
| ccpsecretorigin.blog.epochtimes.com/ | 1 2 3 4 5 6 7 |
| jinxlutianhua.blog.epochtimes.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 |
| lianhuaxiaofo.blog.epochtimes.com/ | 1 2 3 4 5 6 7 8 9 10 |
| zhangtianliang.blog.epochtimes.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 |
| ziyunbokee.blog.epochtimes.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 |
| www.epochweekly.com/ | 1 2 3 4 5 6 7 8 9 10 11 |
| www.facebook.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 |
| zh-cn.facebook.com/ | 1 2 3 4 5 6 7 8 9 |
| fangong.org/ | 1 2 3 |
| www.flickr.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 |
| www.ftchinese.com/ | 1 2 3 4 5 6 7 8 9 10 |

continued on next page...

...continued from previous page

Vanished URLs

| | |
|------------------------------|--|
| code.google.com/ | 1 2 3 4 |
| sites.google.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 |
| www.greatfire.org/ | 1 |
| chinafree.greatzhonghua.org/ | 1 2 3 4 5 |
| shuangzhoukan.hrichina.org/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 |
| www.huffingtonpost.com/ | 1 2 |
| 9jqbqsvg.blog.ifeng.com/ | 1 |
| baike.health.ifeng.com/ | 1 |
| blog.ifeng.com/ | 1 |
| book.ifeng.com/ | 1 |
| finance.ifeng.com/ | 1 2 |
| news.ifeng.com/ | 1 2 3 4 5 6 7 8 |
| tech.ifeng.com/ | 1 |
| v.ifeng.com/ | 1 2 |
| www.imdb.com/ | 1 2 |
| www.laqingdan.net/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 |
| www.linkedin.com/ | 1 2 |
| woeser.middle-way.net/ | 1 2 3 4 5 |
| dzh.mop.com/ | 1 2 3 4 |
| www.my1510.cn/ | 1 2 3 |
| blog.goo.ne.jp/ | 1 2 |
| hk.apple.nextmedia.com/ | 1 2 3 4 5 |
| www1.hk.apple.nextmedia.com/ | 1 |
| cn.nytimes.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 |
| jobmarket.nytimes.com/ | 1 |
| realestateads.nytimes.com/ | 1 |
| www.nytimes.com/ | 1 2 3 4 5 6 7 |
| www.falundafa.org.sg/ | 1 |
| raindog.pixnet.net/ | 1 |
| blog.qq.com/ | 1 |
| cq.qq.com/ | 1 |
| datalib.news.qq.com/ | 1 2 |
| hb.qq.com/ | 1 |
| imgcache.qq.com/ | 1 |
| news.qq.com/ | 1 2 3 4 5 6 |
| user.qzone.qq.com/ | 1 |
| redchinacn.org/ | 1 |
| cn.reuters.com/ | 1 2 3 |
| www.reuters.com/ | 1 |
| www.rfi.fr/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 |
| savedwebhistory.org/ | 1 2 3 |
| m.secretchina.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 |
| gb.home.sina.com/ | 1 |
| home.sina.com/ | 1 2 3 |
| www.slideshare.net/ | 1 2 3 4 5 |
| baike.soso.com/ | 1 2 3 4 5 6 7 8 9 10 |
| news.takungpao.com/ | 1 2 3 |
| www.theepochtimes.com/ | 1 2 3 4 |
| www.theguardian.com/ | 1 |
| www.tiananmenduiuzhi.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 |
| www.tiananmenmother.org/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 |

continued on next page...

...continued from previous page

Vanished URLs

| | |
|-------------------------------|--|
| bbs.tianya.cn/ | 1 2 3 4 5 |
| www.tudou.com/ | 1 2 3 4 |
| blog.udn.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 |
| gb.udn.com/ | 1 2 |
| mag.udn.com/ | 1 2 3 |
| www.verycd.com/ | 1 |
| b5.videopediaworld.com/ | 1 2 3 4 5 6 7 |
| www.voachinese.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 |
| wangjinbo.org/ | 1 2 |
| www.webl.io.jp/ | 1 |
| www.wengewang.org/ | 1 2 3 |
| bbs.wenxuecity.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 |
| cn.uncyclopedia.wikia.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 |
| zh.answers.wikia.com/ | 1 2 3 4 5 6 7 8 9 10 11 |
| commons.wikimedia.org/ | 1 |
| zh.wikinews.org/ | 1 2 3 4 5 6 7 |
| en.wikipedia.org/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 |
| simple.wikipedia.org/ | 1 |
| zh-classical.wikipedia.org/ | 1 |
| zh-yue.wikipedia.org/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 |
| zh.wikipedia.org/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 |
| zh.wikiquote.org/ | 1 2 3 |
| zh.wikisource.org/ | 1 2 3 4 |
| aiyuan.wordpress.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 |
| bannedebook.wordpress.com/ | 1 2 3 4 5 6 7 8 9 10 |
| circlereversed.wordpress.com/ | 1 2 3 4 5 6 7 |
| rightsong.wordpress.com/ | 1 2 3 4 5 6 7 |
| sujiatun.wordpress.com/ | 1 2 3 4 5 6 7 8 9 |
| xijie.wordpress.com/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 |
| hk.knowledge.yahoo.com/ | 1 2 3 4 5 6 7 8 9 10 11 |
| tw.knowledge.yahoo.com/ | 1 2 3 4 5 6 7 8 9 10 11 |
| mymedia.yam.com/ | 1 2 |
| www.ycool.com/ | 1 |
| ycool.com/ | 1 2 |
| v.youku.com/ | 1 2 3 4 5 6 7 8 9 10 11 |
| www.youku.com/ | 1 |
| news.zhengjian.org/ | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 |

Table D.1: Bing's Vanished URLs List

E Bing Censorship Outside China

Below we list a subset of terms which trigger censorship notice for the Chinese version of Bing *outside China*. Readers are welcome to check them out. Remember that the censorship notice may appear on pages after the first. If one does not see the censorship notice within the first five pages, it might be due to changes in Bing's serving system. Nonetheless, we have screenshots of censorship notice for all these terms. Section 7 contains more information.

| | | | | |
|------------------------|------------------------|-------------------------|------------------------|--------------------------|
| 18.dwnnews.com | 18大 18大 | 1989 protest | 1989 天安门 | 2005年5月 |
| 200 活埋 | 2010-2011年突尼 | 2012春晚节目单 | 2012年双普选 | 20周年 |
| 24周年 | 35三十五 | 3月20日兵变 | 5月4日 示威 | 6-4 |
| 64memo | 64事件 | 64 学生运动 | 8964 | 89-64.org |
| 89动乱 | 89 学生运动 | 89学运 | 89运动 | 8八捌eight |
| 9ping | baike.so.com | bajiu | bbs.6park.com | bbs.creaders.net |
| bbs.omnitalk.org | bbs.powerapple.com | bbs.we54.com | bbs.wforum.com | beijing.favolist.com |
| beijingspring.com | bignews | biweekly.hrchina.org | bjzc.org | blog.dwnnews.com |
| blog.wenxuecity.com | bo guagua | book.edoors.com | boxapp.appspot.com | bo yibo |
| Charter 08 | Chen Guangcheng | chinadigitaltimes.net | china.dwnnews.com | chinaism.blogspot.com |
| chinesenewsnet | chinesepen.org | cnjasmine.blogspot.com | cnjasmine.tumblr.com | cnlaf.blogspot.com |
| cnnews.rti.org.tw | cnsoccer.titan24.com | cpajump.centenr.net | culture revolution | dafa |
| dafagood.com | dafahao.com | dailymotion.com | dailynews.sina.com | Dalai |
| Dalai Lama | dawhois.com | deng xiaoping | discuz.club1069.com | dongtaiwang |
| duping.net | ent.6park.com | ent.creaders.net | epaper.epochtimes.com | etienne.clement.im |
| falali | falun | falundafa | falundafa.org | Falun Gong |
| fanglizhi | forum.dwnnews.com | forum.edoors.com | free.21cn.com | freegate |
| freegate download | freenet | freenet-china.org | fyshl.com | gb.udn.com |
| gcd | GlaxoSmithKline | global.dwnnews.com | GONG党 | greatfire.org |
| hbdjk.com | helanonline.cn | henan.kaiwind.com | home.sina.com | hongzhi.li |
| hua guofeng | huanet | huanghuagang.org | hu jintao | human rights |
| hu qili | hyperballad-tibet | intl.feedfury.com | japan | Jasmine |
| ja.wikipedia.org | jiang zemin | jinping | keqiang | library.minghui.org |
| life.dwnnews.com | li keqiang | li peng | li ruihuan | liu si |
| Liusishijian | Liu Xianbin | Liu Xiaobo | liu yunshan | liu四 |
| luntan.epochtimes.com | L月S日 | mag.epochtimes.com | mao zedong | military.dwnnews.com |
| minghui | minghuinews | m.minghui.org | moli | my.opera.com |
| new citizens movement | news.creaders.net | news.edoors.com | news.have8.com | news.hexun.com |
| news.mingpao.com | news.skykiwi.com | news.we54.com | New York Times | nfmz |
| nl.wikipedia.org | observechina | observechina.laogai.org | offshoreleaks.icij.org | ok0.blogspot.com |
| omnitalk.com | opinion.dwnnews.com | package.minghui.org | pastebin.com | peacehall |
| peacehall.com | photo.minghui.org | Princeling | prisoner of the state | pro.17caifu.com |
| qiao shi | Rebiya | shanyuan.vic.mh4u.org | shichang.com | sitetag.us |
| soundofhope.org | sv.wikipedia.org | taiwan.dwnnews.com | talk.news.pts.org.tw | Tiananmen |
| Tiananmen Massacre | Tibet | tibetalk | TOR | Twitter |
| twitter.com | Ultrasurf | us.dongtaiwang.com | uv.gov.blogspot.com | wangdan |
| wang hongwen | wang qishan | wangruowang | wangyang | web.6park.com |
| weekend.minghui.org | wei jianxing | WikiLeaks | woeser.middle-way.net | wu bangguo |
| wuerkaixi | wu guanzheng | www2.renminbao.com | www3.renminbao.com | www3.xinhuanet.com |
| www.64memo.org | www.89-64.org | www.asianews.it | www.backchina.com | www.beijing1989.com |
| www.beyondfirewall.com | www.books.com.tw | www.boxun.com | www.bwsj.hk | www.cdjp.org |
| www.chinacitynews.be | www.chinaelections.org | www.chinagfw.org | www.chinashoes.com | www.chinatimes.com |
| www.chinese.rfi.fr | www.cndev.org | www.cqcounter.com | www.cqlw.com | www.creaders.net |
| www.dailymotion.com | www.docsou.com | www.duping.net | www.dwnnews.com | www.edoors.com |
| www.epochtimes.co.il | www.epochtimes.jp | www.facebook.com | www.falundafa.org | www.financialnews.com.cn |
| www.freechina.net | www.freetibet.org | www.fxingw.com | www.fzxx.org | www.g.dkub.net |
| www.hk20.com | www.hrw.org | www.huanghuagang.org | www.minghui.jp | www.molihua.org |
| www.my1510.cn | www.nandu.com | www.nnmy.com | www.open.com.hk | www.peacehall.com |
| www.qyseo.com | www.remenshipin.net | www.rfi.fr | www.soundofhope.org | www.tibet.org.tw |

continued on next page...

...continued from previous page

| | | | | |
|-------------------|---------------------|----------------|------------------|----------------------|
| www.tsquare.tv | www.wangruowang.org | www.youpai.org | www.yuanming.net | www.zhuichaguoji.org |
| www.zhujiwu.com | www.zueiai.com | www.zuola.com | xhamster.com | xianql.wordpress.com |
| xinshijue.org | xy.sqzw.com | y5w.net | yao wenyuan | yao yilin |
| you行 | yuming.flnet.org | zhang chunqiao | zhao ziyang | zhengsheng |
| zhenshanren | zh.m.wikipedia.org | zhu de | zh.wikipedia.org | zh-yue.wikipedia.org |
| znzx.blogspot.com | 艾未未 | 艾未未工作室 | 安祖缇 | 八九 |
| 八九学运 | 罢工 | 半羽 半羽 | 薄瓜瓜 | 薄熙来 |
| 保福寺 车祸 | 报禁 | 爆炸北京 | 北岛 | 北京 319 |
| 北京 车祸 | 北京枪声 | 表哥 | 不适宜对外公开 | 藏独 |
| 曹建明 | 长安街枪声 | 长老 | 陈光诚 | 陈世炬 |
| 城管 | 程中将 | 持不同政见 | 迟浩田 | 出售手枪 |
| 催泪弹 | 达赖 | 达赖集团 | 达赖喇嘛 | 打倒共党 |
| 打手枪 | 大参考 | 大法 | 大纪元新闻网 | 大家论坛 |
| 大江大海一九四九 | 大中华论坛 | 戴梦得 | 党棍 | 党 魁 |
| 刀 新疆 | 倒台 | 登基 | 邓家贵 | 抵制PX |
| 地下刊物 | 东西南北论坛 | 动乱 | 动网通 | 独裁 |
| 独裁政治 | 独立中文笔会 | 堵截 | 杜导正 | 杜宪 |
| 二十四周年 | 二十周年 | 法lun | 法拉利 | 法十轮 |
| 法滴功 | 翻案 | 反共 | 反日 | 福清纪委爆炸案 |
| 付申奇 | 嘎玛巴 | 改革派 | 高层内斗 | 高官互调 |
| 高志晟 | 蛤蟆 | 公开信 | 公盟法律研究中心 | 共狗 |
| 共和国 | 共同声明 | 顾纪祥 | 官场地震相关 | 硅谷妈妈 |
| 郭泉 | 国际特赦 | 国民教育 | 国务院办公厅 | 海外民运 |
| 汉奸 | 何德普 | 何俊仁 | 何挺 | 河殇 |
| 贺国强 | 洪传 | 洪哲胜 | 洪智 | 红旗飘飘 |
| 胡core | 胡X套 | 胡海峰 | 胡海峰 威视 | 胡平 |
| 胡启立 | 胡温 | 胡温政治 | 胡习 | 华夏文摘 |
| 华语世界论坛 | 黄慈萍 | 黄菊妻子 | 黄子华 | 回良玉 |
| 活埋 名单 | 姬胜德 | 鸡毛信文汇 | 集会 | 己巳月 乙未日 |
| 贾晓烨 | 贾育台 | 建国党 | 姜维平 | 江core |
| 江ZM | 江八点 | 江派 | 江派 江派 | 江猪 |
| 蒋品超 | 讲法 | 酱猪媳 | 截访 | 戒严 |
| 金盾 | 金日成大学 | 金歪日 | 京温 | 经文 |
| 静坐抗议 | 九ping | 九常委 九常委 | 九哥 九哥 | 绝食 |
| 康师傅 儿子 | 康师傅 下架 | 康师傅 中石油 | 抗议 | 劳改 |
| 李洪宽 | 李洪志 | 李继耐 | 李克强 | 李明 小马奔腾 |
| 李现原 | 李源潮 | 李肇星儿媳 | 李卓人 | 炼功 |
| 梁国雄 | 梁振英 简历 | 梁振英 简历 | 良心犯 | 临终遗言 |
| 林彪 | 林瑞雄 | 领导层 | 令jh | 令谷 |
| 刘亚洲 | 流亡 | 六合彩 | 六四 | 六四歌曲 |
| 六四襟章 | 六四事件 | 六四学潮 | 六四学运 | 六四运动 |
| 龙应台 | 路透 | 陆肆事件 | 轮十功 | 盲人律师 |
| 美国之音 | 蒙独 | 民主女神 | 民主中国 | 明慧 |
| 茉莉花事件 | 穆斯林 | 南方都市报 | 南周事件 | 南周 新年献词 |
| 囊帕拉枪杀事件 | 闹事 | 内斗 | 内蒙古抗议 | 内争人权 |
| 念念不忘 | 纽约时报 | 潘国平 | 彭 iphone | 彭 手机 |
| 平反64 | 平反六四 | 平西王 | 迫害 | 普选 |
| 七一 上街 | 秦城监狱 | 秦裕 | 权斗 | 权力 交接 |
| 燃烧瓶 | 热比娅 | 赛风 | 三角政治同盟 | 删除 负面信息 |
| 上海孤儿院 | 尚福林 辞职 | 师涛 | 史记 温秀秀 | 世界通 |
| 事件 319 | 示威 | 首长 | 水扁 | 四川朱昱 |
| 四人帮 | 四 事件 | 台海 | 台湾狗 | 台湾建国运动组织 |
| 太子党 | 谭作人 | 唐柏桥 | 讨伐 | 天安门事件 |
| 天安门一代 | 天安 事件 | 天鹅绒 | 天鹅绒行动 | 天怒 |
| 天上鬼 | 天神右翼 | 天生我材必有用 | 天线宝宝 方便面 | 天葬 |

continued on next page...

...continued from previous page

| | | | | |
|----------|-------------|---------|--------|----------|
| 透视中国 | 推特 | 退党潮 | 退共党 | 虔震 |
| 万维读者论坛 | 汪岷 | 王qs | 王宝森 | 王炳章 |
| 王丹 | 王沪宁 | 王立军 | 王若望 | 王文怡 |
| 王希哲 | 王兆国 | 威视腐败案 | 唯色 | 维权人士 |
| 未亡人 | 温JB | 温如春 | 温爷爷 | 温影帝 |
| 温云松 | 瘟加饱 | 瘟疫 | 文炎 | 乌尔 |
| 污黑 | 吾尔开希 | 无界浏览器 | 西藏问题 | 西单民主墙 |
| 习干平 | 习和平 | 习假新闻 | 习近平 | 习明泽 |
| 习仲勋 | 下台 | 先皇 | 香港泛民主派 | 香港民主 |
| 小胡 | 邪教 | 新疆暴力 | 新疆恐怖 | 新唐人 |
| 新闻封锁 | 新闻自由 | 徐邦泰 | 徐才厚 | 徐水良 |
| 许万平 | 学运 | 亚洲周刊 | 殷虹 | 延安日记 |
| 央视内部晚会 | 姚月谦 | 叶剑英女婿 | 一党制 | 伊力哈木 呼吁 |
| 伊宁事变 | 异议人士 | 婴儿煲汤 | 影帝 wen | 拥护 薄 |
| 游行 | 右派 | 于幼军 | 余杰 | 余秋雨 |
| 余姚 示威 | 冤民大同盟 | 冤罪 | 袁冬 | 袁红冰 |
| 造反 | 占领青岛 | 占领乌鲁木齐 | 占领西宁 | 张德江 |
| 张高丽 | 张立昌 | 章沁生 | 朝鲜 政变 | 朝阳医院 骆家辉 |
| 朝阳医院自行离开 | 赵本山退出春晚 | 赵品潞 | 赵天韶 | 赵紫阳 |
| 真理部 | 整肃 | 政治局常委内幕 | 政治局会议 | 政治内幕 |
| 正腐 | 正邪大决战 | 支持 薄 | 中国民主党 | 中国民主正义党 |
| 中国新民党 | 中华民国建国102周年 | 中南海斗争 | 中南海门保镖 | 中央警卫局 |
| 重庆万盛事件 | 周YK | 周薄 | 周本顺 | 周康康 |
| 周晓川 | 周永康 | 周正毅 | 朱令案 | 祝均一 |
| 专政 | 转世 | 自由之门 | 总理 影帝 | |

Table E.1: Terms triggering censorship notice outside China